

# Phase estimation without a priori knowledge in the presence of loss

Jan Kołodyński<sup>1</sup> and Rafał Demkowicz-Dobrzański<sup>1</sup>

<sup>1</sup>*Institute of Theoretical Physics, University of Warsaw, ul. Hoża 69, PL-00-681 Warszawa, Poland*

We find the optimal scheme for quantum phase estimation in the presence of loss when no a priori knowledge on the estimated phase is available. We prove analytically an explicit lower bound on estimation uncertainty, which shows that, as a function of number of probes, quantum precision enhancement amounts at most to a constant factor improvement over classical strategies.

PACS numbers: 03.65.Ta, 06.20.Dk, 42.50.St

## I. INTRODUCTION

Owing to highly promising predictions of the theory of precise quantum measurements and parameter estimation, as well as significant progress in quantum state engineering, the task of phase shift determination has recently been readdressed both theoretically and experimentally [1–11]. In classical systems the precision of the estimated phase scales with the amount of available resources as  $1/\sqrt{N}$ , the so called *Standard Quantum Limit* (SQL) or more commonly the “shot noise”. Traditionally,  $N$  denotes the number of independent measuring probes, repetitions or copies of a system. The potential precision boost offered by quantum mechanics stems from the possibility of preparing  $N$  copies of a system in a highly entangled state, particularly sensitive to the variations of the estimated parameter [1–3]. In ideal scenarios, these states yield phase estimation precision which scales as  $1/N$  and is referred to as the *Heisenberg Limit* (HL).

Environmentally induced decoherence, however, significantly affects the performance of entanglement based quantum strategies [12–21] with photon loss being its most relevant source in optical implementations. The need to balance the phase sensitivity and robustness against losses results in states performing better than SQL yet falling short of HL [18, 19]. Other approaches, trying to mimic the quantum enhanced strategies using multiple-pass technique [8] are even more susceptible to losses and cannot compete with the optimally designed entangled states [22]. Despite the quantitative improvement of precision offered by quantum states in the presence of loss, it has remained an unsolved problem whether in the asymptotic regime  $N \rightarrow \infty$  quantum states offer better than SQL scaling, i.e.,  $c/N^\alpha$  with  $\alpha > 1/2$ .

In this paper we solve the problem of optimal phase estimation in the presence of loss with no a priori knowledge, and prove analytically that even for arbitrarily small loss, quantum enhancement does not offer better than  $c/\sqrt{N}$  scaling for  $N \rightarrow \infty$ , and

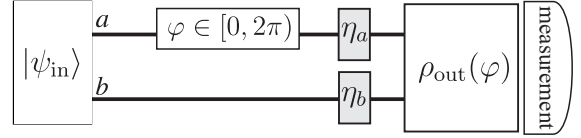


Figure 1. Phase estimation setup. Channel  $a$  acquires a phase  $\varphi$  relative to channel  $b$ . Losses are modeled by two beam splitters with power transmissions  $\eta_a, \eta_b$ .

the only gain over classical strategies is a smaller multiplicative constant  $c$ . It should be emphasized that the proof contains the most general description of a quantum measurement, hence its conclusions are valid also for adaptive schemes (see Appendix C), which are especially interesting from a practical point of view [23, 24].

## II. MODEL

Two approaches to phase estimation are typically pursued. In the first, *local approach*, a measurement scheme is devised, which offers the highest sensitivity to phase deviations from an a priori known value,  $\varphi = \varphi_0$ . This is achieved by finding a strategy that maximizes the *quantum Fisher information*,  $F_Q$ , which defines the lower bound on the precision of the estimated phase through  $\delta\varphi \geq 1/\sqrt{F_Q}$  [25–28]. The optimal states have been found both for lossless [1, 2] (the so called N00N states) and more realistic lossy scenarios [18, 19].

The second approach, which we will pursue in this paper and refer to as the *global approach*, assumes *no a priori knowledge* about the phase, so that  $\varphi$  is equiprobably distributed over the  $[0, 2\pi)$  region. We consider a general pure  $N$  photon two-mode state [29]

$$|\psi_{\text{in}}\rangle = \sum_{n=0}^N \alpha_n |n, N-n\rangle, \quad (1)$$

which is fed into an interferometer with a relative phase delay  $\varphi$  (see Fig. 1). Apart from acquiring the phase via the unitary  $U_\varphi = e^{-i\varphi a^\dagger a}$ , the state experiences losses modeled by two beam splitters with power transmissions  $\eta_a$  and  $\eta_b$  [30]. The output state then takes the form  $\rho_{\text{out}}(\varphi) = U_\varphi \rho_{\text{out}} U_\varphi^\dagger$ , where

$$\rho_{\text{out}} = \sum_{l_a=0}^N \sum_{l_b=0}^{N-l_a} |\phi^{l_a, l_b}\rangle \langle \phi^{l_a, l_b}|, \quad (2)$$

with subnormalized conditional states corresponding to  $l_a$  and  $l_b$  photons lost in arms  $a$  and  $b$  respectively

$$|\phi^{l_a, l_b}\rangle = \sum_{n=l_a}^{N-l_b} \alpha_n \beta_n^{l_a, l_b} |n-l_a, N-n-l_b\rangle \quad (3)$$

where

$$\beta_n^{l_a, l_b} = \sqrt{B_{l_a}^n(\eta_a) B_{l_b}^{N-n}(\eta_b)}, \quad B_l^n(\eta) = \binom{n}{l} (1-\eta)^l \eta^{n-l}. \quad (4)$$

Keeping the reasoning most general, the information about  $\varphi$  is extracted via a measurement on  $\rho_{\text{out}}(\varphi)$  described by a Positive Operator Valued Measure (POVM),  $\{M_r\}$ ,  $\sum_r M_r = \mathbb{1}$ . The outcome  $r$  is observed with probability  $p(r|\varphi) = \text{Tr}\{\rho_{\text{out}}(\varphi) M_r\}$ , and the estimated phase inferred from it is defined by an estimator  $\tilde{\varphi}(r)$ . Optimization procedure with respect to a given cost function  $C(\varphi, \tilde{\varphi})$  amounts to finding the state  $|\psi\rangle$ , the measurement  $\{M_r\}$ , and the estimator  $\tilde{\varphi}(r)$  that minimize the cost function averaged over a flat a priori phase distribution

$$\langle C \rangle = \int \frac{d\varphi}{2\pi} \sum_r p(r|\varphi) C(\varphi, \tilde{\varphi}(r)). \quad (5)$$

Let  $C(\varphi, \tilde{\varphi}) = C(\varphi - \tilde{\varphi}) = \sum_{n=-\infty}^{\infty} c_n e^{in(\varphi - \tilde{\varphi})}$ , be an arbitrary real symmetric cost function ( $c_n = c_{-n} \leq 0$  for  $n \neq 0$ ) respecting the cyclic nature of  $\varphi$  [26, 31].

### III. OPTIMIZATION

Thanks to the flat a priori phase distribution, the problem enjoys a symmetry with respect to an arbitrary phase shift  $U_\varphi$ . The search for the optimal measurement strategy may be restricted to the class of covariant POVM  $\{M_{\tilde{\varphi}}\}$  [26, 31, 32] parameterized by a continuous parameter  $\tilde{\varphi}$ :  $M_{\tilde{\varphi}} = U_{\tilde{\varphi}} \Xi U_{\tilde{\varphi}}^\dagger$ , where  $\Xi$  is a positive semi-definite operator satisfying the POVM completeness constraint  $\int \frac{d\tilde{\varphi}}{2\pi} U_{\tilde{\varphi}} \Xi U_{\tilde{\varphi}}^\dagger = \mathbb{1}$ . With the above substitution, the average cost function simplifies to

$$\langle C \rangle = \int \frac{d\varphi}{2\pi} \text{Tr}\{\rho_{\text{out}}(\varphi) \Xi\} C(\varphi) \quad (6)$$

and  $\langle C \rangle$  has to be minimized only over the choice of the input state  $|\psi\rangle_{\text{in}}$  and the seed operator  $\Xi$ .

In order to find the optimal  $\Xi$ , one can rewrite Eq. (2) in the form  $\rho_{\text{out}} = \bigoplus_{N'=0}^N \rho_{\text{out}}^{N'}$ , with  $\rho_{\text{out}}^{N'} = \sum_{l_a=0}^{N-N'} |\phi^{l_a, N-N'-l_a}\rangle \langle \phi^{l_a, N-N'-l_a}|$ , which reveals the block structure with respect to the total number of surviving photons  $N'$ . Therefore, without loss of generality, we may impose an analogous block structure on the seed operator  $\Xi = \bigoplus_{N'=0}^N \Xi^{N'}$ . Physically, such a block structure implies that a non-demolition photon number measurement had been performed at the output, before any further phase measurements have taken place. Following the reasoning presented in [26, 31] it can be shown that without loosing optimality, the input state parameters  $\alpha_n$  can be chosen real, in which case the optimal seed operator  $\Xi_{\text{opt}}^{N'} = |e_{N'}\rangle \langle e_{N'}|$ , where  $|e_{N'}\rangle = \sum_{n=0}^{N'} |n, N'-n\rangle$  (see Appendix A).

In what follows we choose the cost function  $C(\varphi - \tilde{\varphi}) = 4 \sin^2 \frac{\varphi - \tilde{\varphi}}{2}$  ( $c_0 = 2, c_1 = c_{-1} = -1$ ) and denote its average by  $\widetilde{\delta^2 \varphi}$ , as it is the simplest cost function approximating the variance for narrow distributions [3].

Performing the integration in Eq. (6) the average cost function reads:

$$\widetilde{\delta^2 \varphi} = 2 - \boldsymbol{\alpha}^\dagger \mathbf{A} \boldsymbol{\alpha}, \quad (7)$$

where non-zero elements of the matrix  $\mathbf{A}$  read:

$$A_{n-1, n} = A_{n, n-1} = \sum_{l_a, l_b=0}^{n, N-n} \beta_n^{l_a, l_b} \beta_{n-1}^{l_a, l_b}. \quad (8)$$

Hence, the minimal cost equals  $\widetilde{\delta^2 \varphi} = 2 - \lambda_{\text{max}}$ , where  $\lambda_{\text{max}}$  is the maximal eigenvalue of the matrix  $\mathbf{A}$ , and the corresponding eigenvector provides the optimal input state parameters  $\boldsymbol{\alpha}$ .

#### A. Numerical solution

Numerical results of the above eigenvalue problem are presented in Fig. 2. Black lines depict phase estimation uncertainty  $\delta\varphi$  of the optimal quantum strategy plotted as a function of  $N$  for  $\eta_a = \eta_b \in \{0.6, 0.8, 1\}$ . In the absence of loss the optimal quantum curve tends to the Heisenberg scaling, whereas, when losses are present, it flattens significantly with increasing  $N$ . The inset depicts the form of the optimal state. With increasing degree of loss the distribution of  $\alpha_n$  for the optimal state becomes more peaked as compared with the lossless

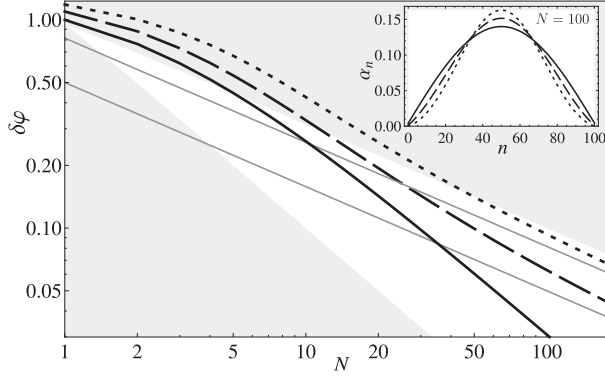


Figure 2. Log-log plot of optimal phase estimation uncertainty as a function of number of photons used for three different levels of loss (equal in both arms):  $\eta = 1$  (solid),  $\eta = 0.8$  (dashed),  $\eta = 0.6$  (dotted). White area in the middle of the picture corresponds to  $1/N < \delta\varphi < 1/\sqrt{N}$ . Gray lines represent asymptotic bounds given by Eq. (12) for  $\eta = 0.8$ ,  $\eta = 0.6$ . The inset depicts the structure of the optimal states for the three levels of loss for  $N = 100$ .

case  $\alpha_n = \sqrt{\frac{2}{N+2}} \sin\left[\frac{(n+1)\pi}{N+2}\right]$  [3]. This behavior can be intuitively understood in a similar fashion as in the local approach [18, 19], where the  $N00N$  states with only two non-zero coefficients  $\alpha_0, \alpha_N$ , are the most sensitive to the phase shift but extremely vulnerable to loss. In the presence of loss, larger weights need to be ascribed to intermediate coefficients, in order to preserve quantum superposition even after some photons are lost. The same effect of increasing weights of intermediate coefficients at the expense of marginal ones is also present in the global approach.

### B. Asymptotic bounds

We now move on to present the main result of the paper. Numerical results presented above and the ones obtained within the local approach [18, 19] indicate that in the presence of loss, phase estimation uncertainty  $\delta\varphi$  departs from the HL and asymptotically approaches  $c/\sqrt{N}$ . Until now, however, an analytical proof of the above conjecture was missing.

Let us first derive an upper bound on the maximal eigenvalue  $\lambda_{\max}$  of matrix  $\mathbf{A}$  in Eq. (7). Without loss of generality, we assume that  $\eta_a \leq \eta_b$ . Clearly, setting  $\eta_b = 1$  can only improve our estimation—hence  $\lambda_{\max}$  increases. For  $\eta_a = \eta < 1$ ,  $\eta_b = 1$  the nonzero matrix elements read:  $A_{n,n-1} = \sum_{l=0}^n \sqrt{B_l^n(\eta) B_l^{n-1}(\eta)}$ .

Recall that for an arbitrary normalized vector  $\mathbf{v}$ ,

$\mathbf{v}^\dagger \mathbf{A} \mathbf{v} \leq \lambda_{\max}$ . Let  $\boldsymbol{\alpha}$  be the eigenvector corresponding to  $\lambda_{\max}$ :  $\boldsymbol{\alpha}^\dagger \mathbf{A} \boldsymbol{\alpha} = \lambda_{\max}$ . The fact that all matrix elements of  $\mathbf{A}$  are non-negative, implies  $\forall_n \alpha_n \geq 0$ .

Let us now define a matrix  $\mathbf{A}'$ , such that all nonzero entries of  $\mathbf{A}$  are replaced by the maximum matrix element  $A^\dagger = \max_n \{A_{n,n-1}\} = A_{N,N-1}$ . Since  $\alpha_n \geq 0$  and  $A'_{n,m} \geq A_{n,m} \geq 0$  we can write:

$$\lambda_{\max} = \boldsymbol{\alpha}^\dagger \mathbf{A} \boldsymbol{\alpha} \leq \boldsymbol{\alpha}^\dagger \mathbf{A}' \boldsymbol{\alpha} \leq \lambda'_{\max}, \quad (9)$$

where  $\lambda'_{\max}$  is the maximal eigenvalue of  $\mathbf{A}'$ .  $\lambda'_{\max}$  can be found analytically by noting the following recurrence relation for the characteristic polynomial of  $\mathbf{A}'$ :  $\det \boldsymbol{\Lambda}_{n+1} = -\lambda \det \boldsymbol{\Lambda}_n - A^{\dagger 2} \det \boldsymbol{\Lambda}_{n-1}$ , where  $\boldsymbol{\Lambda} = \mathbf{A}' - \lambda \mathbb{1}$ , while  $\boldsymbol{\Lambda}_n$  are  $(n+1) \times (n+1)$  submatrices of  $\boldsymbol{\Lambda}$ . The solution of the recurrence relation reads  $\det(\boldsymbol{\Lambda}) = D_{N+1}(-\lambda, A^{\dagger 2})$ , where  $D_n(x, a^2) = a^n \frac{\sin[(n+1)\arccos(\frac{x}{2a})]}{\sin[\arccos(\frac{x}{2a})]}$  is the Dickson polynomial [33] of the  $n$ th order. The largest eigenvalue corresponds to the largest root of  $\det(\boldsymbol{\Lambda})$ ,  $\lambda'_{\max} = 2A^\dagger \cos\left[\frac{\pi}{(N+2)}\right]$ .

We can finally write explicitly the lower bound on the variance:

$$\widetilde{\delta^2\varphi} \geq 2 \left[ 1 - \cos\left(\frac{\pi}{N+2}\right) \sum_{l=0}^N \sqrt{B_l^N(\eta) B_l^{N-1}(\eta)} \right]. \quad (10)$$

Expanding the above formula in the limit  $N \rightarrow \infty$  we get:

$$\widetilde{\delta^2\varphi} \geq \frac{1-\eta}{4\eta N} + O\left(\frac{1}{N^2}\right), \quad (11)$$

which proves that for  $\eta < 1$ ,  $\delta\varphi$  scales as  $c/\sqrt{N}$ .

A tighter bound can be analogously derived for the case  $\eta_a = \eta_b = \eta$ , by noting that  $\max_n \{A_{n,n-1}\} = A_{\lceil \frac{N}{2} \rceil, \lceil \frac{N}{2} \rceil - 1}$ . In the limit  $N \rightarrow \infty$  we get:

$$\widetilde{\delta^2\varphi} \geq \frac{1-\eta}{\eta N} + O\left(\frac{1}{N^2}\right). \quad (12)$$

### C. Optimal classical strategy

For the sake of comparison, we also derive the optimal classical phase estimation strategy, in which a coherent state with mean photon number  $N$  is sent to an initial beam splitter of transmissivity  $\tau_{\text{in}}$ , whose output feeds paths  $a$  and  $b$  of the interferometer. We assume no additional external phase reference, hence the state is effectively a mixture of terms with

a different total photon number. The optimal seed POVM is  $\bigoplus_{N'=0}^{\infty} \Xi_{\text{opt}}^{N'}$  yielding:

$$\widetilde{\delta^2\varphi} = 2 - \frac{2\mathcal{B}[N\eta_a\tau_{\text{in}}]\mathcal{B}[N\eta_b(1-\tau_{\text{in}})]}{N\sqrt{\eta_a\tau_{\text{in}}\eta_b(1-\tau_{\text{in}})}}. \quad (13)$$

where  $\mathcal{B}(x) = e^{-x} \sum_{n=0}^{\infty} \frac{x^n}{n!} \sqrt{n}$  is the Bell polynomial of order 1/2. For strong beams ( $N \rightarrow \infty$ ) up to the first order in  $1/N$ ,  $\widetilde{\delta^2\varphi} \approx \left( \frac{1}{\tau_{\text{in}}\eta_a} + \frac{1}{(1-\tau_{\text{in}})\eta_b} \right) / 4N$  and is minimized for the choice  $\tau_{\text{in}} = 1/(1 + \sqrt{\eta_a/\eta_b})$

$$\widetilde{\delta^2\varphi} \approx \frac{1}{4N} \left( \frac{1}{\sqrt{\eta_a}} + \frac{1}{\sqrt{\eta_b}} \right)^2, \quad (14)$$

which is exactly the same formula as for the optimal classical strategy in the local approach [19].

#### IV. CONCLUSIONS

Results presented in the paper indicate that, while quantum enhanced protocols provide quantitative boost in the estimation precision, the presence of loss unavoidably causes the precision scaling to become classical in the limit of large number of resources  $N$ . The asymptotic gain of quantum enhanced protocols

amounts just to a smaller multiplicative constant  $c$  in the scaling law  $c/\sqrt{N}$ . Comparing Eq. (14) (with  $\eta_a = \eta$ ,  $\eta_b = 1$ ) with the bound given in Eq. (11) we may conclude that asymptotically quantum enhanced protocols provide at most a factor of

$$\lim_{N \rightarrow \infty} \frac{\delta\varphi^{\text{classical}}}{\delta\varphi^{\text{quantum}}} \leq \sqrt{\frac{1 + \sqrt{\eta}}{1 - \sqrt{\eta}}} \quad (15)$$

decrease in the uncertainty of estimation. In the case  $\eta_a = \eta_b = \eta$ , using a tighter bound (12) the above factor reads  $1/\sqrt{1-\eta}$ . We conjecture that the fact that losses necessarily turn HL into  $c/\sqrt{N}$  is a general feature of all quantum estimation problems, such as estimation of direction, Cartesian frames etc.

#### ACKNOWLEDGMENTS

We acknowledge many fruitful discussions with Konrad Banaszek. This research was supported by the European Commission under the Integrating Project Q-ESSENCE and the Foundation for Polish Science under the TEAM program.

After this work has been completed, analogous conclusions have been presented within the complementary local approach [34].

#### Appendix A: Optimal measurement

Substituting the output state  $\rho_{\text{out}} = \bigoplus_{N'=0}^N \rho_{\text{out}}^{N'}$  and the seed operator  $\Xi = \bigoplus_{N'=0}^N \Xi^{N'}$  to Eq. (6), we get an explicit formula for the average cost function:

$$\langle C \rangle = \sum_{N'=0}^N \sum_{l_a=0}^{N-N'} \sum_{n,m=l_a}^{N'+l_a} C_{nm} \beta_n^{l_a, l_b} \beta_m^{l_a, l_b} \alpha_n^* \alpha_m \Xi_{n-l_a, m-l_a}^{N'}, \quad (A1)$$

where  $l_b = N - N' - l_a$ ,  $C_{nm} = \int \frac{d\varphi}{2\pi} C(\varphi) e^{i(n-m)\varphi}$  and  $\Xi_{n', m'}^{N'} = \langle n', N' - n' | \Xi^{N'} | m', N' - m' \rangle$ . The completeness constraint  $\int \frac{d\varphi}{2\pi} U_{\varphi} \Xi U_{\varphi}^{\dagger} = \mathbb{1}$  implies that  $\Xi_{n', n'}^{N'} = 1$ . Therefore, if restricted to  $m = n$  terms, the sum (A1) reduces to a constant  $c_0 = C_{00}$ . Changing the summation order we can rewrite Eq. (A1) as

$$\langle C \rangle - c_0 = \sum_{\substack{n,m=0 \\ n \neq m}}^N \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} C_{nm} \beta_n^{l_a, l_b} \beta_m^{l_a, l_b} \alpha_n^* \alpha_m \Xi_{n-l_a, m-l_a}^{N-l_a-l_b}. \quad (A2)$$

Now, as for all  $n \neq m$  cost coefficients  $C_{nm} \leq 0$ , we get the following lower bound on the average cost

$$\langle C \rangle - c_0 \geq \sum_{\substack{n,m=0 \\ n \neq m}}^N \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} C_{nm} \beta_n^{l_a, l_b} \beta_m^{l_a, l_b} |\alpha_n^*| |\alpha_m| \left| \Xi_{n-l_a, m-l_b}^{l_a+l_b} \right| \quad (\text{A3})$$

$$\geq \sum_{\substack{n,m=0 \\ n \neq m}}^N \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} C_{nm} \beta_n^{l_a, l_b} \beta_m^{l_a, l_b} |\alpha_n^*| |\alpha_m|. \quad (\text{A4})$$

The first inequality is saturated by choosing input state's and seed operator's coefficients to be real. The second inequality follows from  $\Xi_{n',m'}^{N'} \leq \sqrt{\Xi_{m',m'}^{N'} \Xi_{n',n'}^{N'}} = 1$ , which is a consequence of positive semi-definiteness of  $\Xi^{N'}$  and the completeness constraint. Both inequalities are saturated for  $\Xi_{\text{opt}}^{N'} = |e_{N'}\rangle \langle e_{N'}|$ , where  $|e_{N'}\rangle = \sum_{n=0}^{N'} |n, N' - n\rangle$ . This proves the optimality of the measurement considered in the paper.

## Appendix B: Distinguishability of photons

If photons traveling through the interferometer are distinguishable, e.g. they are prepared in different time bins, the dimension of the Hilbert space needed to describe the state of  $N$  photons is  $2^N$ , as opposed to  $N+1$  for the indistinguishable case. In fact, the indistinguishable case may be considered as a restriction of the former space to its fully symmetric subspace. We prove below that considering distinguishable photons is of no use, since the optimality can always be attained within the class of states belonging to the fully symmetric (bosonic) subspace. Let

$$|\psi_N\rangle = \sum_{\mathbf{n}=0^N}^{\mathbf{1}^N} \alpha_{\mathbf{n}} |\mathbf{n}\rangle, \quad (\text{B1})$$

be a general state of  $N$  distinguishable photons traveling through the interferometer, where the sum runs over all  $N$ -bit sequences  $\mathbf{n}$ , with  $|\mathbf{n}\rangle = |n_1\rangle \otimes \dots \otimes |n_N\rangle$ , where  $|n_i\rangle = |1\rangle$  ( $|0\rangle$ ) denotes a photon in the  $i$ th time bin, propagating in the  $a$  ( $b$ ) arm of the interferometer respectively.

Taking loss into account, we additionally need to track the time slots in which photons were lost. We define a binary string  $\mathbf{l}_a = l_{a,1} l_{a,2} \dots l_{a,N}$  with 1s representing the time bins in which photon was lost in arm  $a$  and similarly  $\mathbf{l}_b$  for the arm  $b$ . The general seed operator has a block diagonal structure with respect to different patterns of surviving photons:  $\Xi = \bigoplus_{\mathbf{N}'=0^N}^{\mathbf{1}^N} \Xi^{N'}$ , where 1s in the binary string  $\mathbf{N}'$  denote the time bins in which photons were successfully transmitted. Formally, using bitwise subtraction, we can write  $\mathbf{N}' = \mathbf{1} - \mathbf{l}_a - \mathbf{l}_b$ . Written in a basis  $\Xi^{N'} = \sum_{\mathbf{n}', \mathbf{m}'=0^{\mathbf{N}'}}^{\mathbf{1}^{\mathbf{N}'}} \Xi_{\mathbf{n}', \mathbf{m}'}^{N'} |\mathbf{n}'\rangle \langle \mathbf{m}'|$ , in which  $\mathbf{n}'$  stands for a string with  $N'$  bits placed at positions corresponding to 1s in  $\mathbf{N}'$  with complementary positions left empty (neither 0 nor 1). In order to simplify the notation, for any binary sequence  $\mathbf{x}$ , we denote by  $x = |\mathbf{x}|$  the number of 1s in the sequence. Moreover, we use a notation  $\mathbf{x} \setminus \mathbf{y}$  for a binary string  $\mathbf{x}$  with empty entries at positions corresponding to 1s in  $\mathbf{y}$ .

Adapting Eq. (A2) to the distinguishable photon case, we get:

$$\langle C \rangle - c_0 = \sum_{\substack{\mathbf{n}, \mathbf{m}=0 \\ \mathbf{n} \neq \mathbf{m}}}^{\mathbf{1}} \sum_{\mathbf{l}_a=0}^{\min(\mathbf{n}, \mathbf{m})} \sum_{\mathbf{l}_b=0}^{\mathbf{1}-\max(\mathbf{n}, \mathbf{m})} C_{nm} \gamma_n^{l_a, l_b} \gamma_m^{l_a, l_b} \alpha_n^* \alpha_m \Xi_{\mathbf{n} \setminus (\mathbf{l}_a + \mathbf{l}_b), \mathbf{m} \setminus (\mathbf{l}_a + \mathbf{l}_b)}^{\mathbf{1} - (\mathbf{l}_a + \mathbf{l}_b)} \quad (\text{B2})$$

where min, max should be understood as bitwise operations,  $\gamma_n^{l_a, l_b} = \sqrt{(1 - \eta_a)^{l_a} \eta_a^{n-l_a} (1 - \eta_b)^{l_b} \eta_b^{N-n-l_b}}$  and for simplicity we have put  $\mathbf{0} = \mathbf{0}^N$ ,  $\mathbf{1} = \mathbf{1}^N$ .

We now split the sums over  $\mathbf{l}_i$  into sum over  $l_i$  (number of 1s in  $\mathbf{l}_i$ ) and the sum over permutation of 1s within  $\mathbf{l}_i$ . We proceed analogously for summations over  $\mathbf{n}$  ( $\mathbf{m}$ ) obtaining

$$\begin{aligned} \langle C \rangle - c_0 &= \sum_{\substack{n,m=0 \\ n \neq m}}^N \sum_{\substack{l_a=0 \\ l_b=0}}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} C_{nm} \gamma_n^{l_a, l_b} \gamma_m^{l_a, l_b} \\ &= \sum_{\substack{\mathbf{n}=\mathbf{0} \\ |\mathbf{n}|=n}}^1 \sum_{\substack{\mathbf{m}=\mathbf{0} \\ |\mathbf{m}|=m}}^1 \alpha_n^* \alpha_m \sum_{\substack{l_a=0 \\ |\mathbf{l}_a|=l_a}}^{\min(\mathbf{n}, \mathbf{m})} \sum_{\substack{l_b=0 \\ |\mathbf{l}_b|=l_b}}^{1-\max(\mathbf{n}, \mathbf{m})} \Xi_{\mathbf{n} \setminus (l_a+l_b), \mathbf{m} \setminus (l_a+l_b)}^{1-(l_a+l_b)} \end{aligned} \quad (\text{B3})$$

In order to proceed further let us for the moment specialize to lossless case  $\eta_a = \eta_b = 1$ , where the above formula simplifies to:

$$\langle C \rangle - c_0 = \sum_{\substack{n,m=0 \\ n \neq m}}^N C_{nm} \sum_{\substack{\mathbf{n}=\mathbf{0} \\ |\mathbf{n}|=n}}^1 \sum_{\substack{\mathbf{m}=\mathbf{0} \\ |\mathbf{m}|=m}}^1 \alpha_n^* \alpha_m \Xi_{\mathbf{n}, \mathbf{m}}^1. \quad (\text{B4})$$

$\Xi$  needs to be a positive semi-definite operator, and by completeness constraint  $\Xi_{\mathbf{m}, \mathbf{n}} = \delta_{\mathbf{m}, \mathbf{n}}$ , whenever  $n = m$ . Since diagonal blocks of  $\Xi$  (corresponding to  $n = m$ ) are proportional to identity, it implies that none of the off-diagonal blocks of  $\Xi$  (corresponding to  $n \neq m$ ) can have a singular value larger than 1. This can be proven as follows. Let us assume that for certain block  $(m, n)$  ( $n \neq m$ ), the largest singular value  $\lambda > 1$ , and let  $|\mathbf{v}_m\rangle, |\mathbf{w}_n\rangle$  be the normalized left and right singular vectors corresponding to singular value  $\lambda$ ,  $|\mathbf{v}_m| = m$ ,  $|\mathbf{w}_n| = n$ . Defining  $|\mathbf{z}\rangle = |\mathbf{v}_m\rangle - |\mathbf{w}_n\rangle$ , we calculate

$$\langle \mathbf{z} | \Xi | \mathbf{z} \rangle = \langle \mathbf{v}_n | \Xi | \mathbf{v}_n \rangle + \langle \mathbf{w}_m | \Xi | \mathbf{w}_m \rangle - 2\Re \langle \mathbf{v}_n | \Xi | \mathbf{w}_m \rangle = 2(1 - \lambda) < 0, \quad (\text{B5})$$

which contradicts the positivity semi-definiteness of  $\Xi$ . Because all singular values of any  $(n, m)$  block of  $\Xi$  are smaller than one, the following inequality holds:  $\sum_{\substack{\mathbf{n}=\mathbf{0} \\ |\mathbf{n}|=n}}^1 \sum_{\substack{\mathbf{m}=\mathbf{0} \\ |\mathbf{m}|=m}}^1 \alpha_n^* \alpha_m \Xi_{\mathbf{n}, \mathbf{m}} \leq \alpha_n^* \alpha_m$ ,  $\alpha_n = \sqrt{\sum_{\substack{\mathbf{n}=\mathbf{0} \\ |\mathbf{n}|=n}}^1 |\alpha_n|^2}$ .

This leads to a bound on the cost function in the lossless case

$$\langle C \rangle - c_0 \geq \sum_{\substack{n,m=0 \\ n \neq m}}^N C_{nm} \alpha_n \alpha_m^*, \quad (\text{B6})$$

proving that one can achieve optimality restricting oneself to indistinguishable photons.

Returning to Eq. (B4), we see that we can apply a similar argumentation making use of positive semi-definiteness of  $\Xi_{\mathbf{m}, \mathbf{n}}^{(l_a, l_b)} = \sum_{\substack{l_a=0 \\ |\mathbf{l}_a|=l_a}}^{\min(\mathbf{n}, \mathbf{m})} \sum_{\substack{l_b=0 \\ |\mathbf{l}_b|=l_b}}^{1-\max(\mathbf{n}, \mathbf{m})} \Xi_{\mathbf{m} \setminus (l_a+l_b), \mathbf{n} \setminus (l_a+l_b)}^{1-(l_a+l_b)}$  operator. We notice that the completeness constraint again implies a block structure of  $\Xi^{(l_a, l_b)}$  with respect to  $m = |\mathbf{m}|, n = |\mathbf{n}|$ , with diagonal elements of diagonal blocks  $(n, n)$  being now  $\sum_{\substack{l_a=0 \\ |\mathbf{l}_a|=l_a}}^n \sum_{\substack{l_b=0 \\ |\mathbf{l}_b|=l_b}}^{1-n} 1 = \binom{n}{l_a} \binom{N-n}{l_b}$ . This implies that the maximum singular value of any  $(m, n)$  block of  $\Xi^{(l_a, l_b)}$  is constrained by  $\binom{\min(n, m)}{l_a} \binom{N-\max(n, m)}{l_b}$ . As a result, we obtain the following bound:

$$\begin{aligned} \langle C \rangle - c_0 &\geq \sum_{\substack{n,m=0 \\ n \neq m}}^N \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} C_{nm} \gamma_n^{l_a, l_b} \gamma_m^{l_a, l_b} \binom{\min(n, m)}{l_a} \binom{N-\max(n, m)}{l_b} |\alpha_n| |\alpha_m^*| \\ &\quad \sum_{\substack{n,m=0 \\ n \neq m}}^N \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} C_{nm} \gamma_n^{l_a, l_b} \gamma_m^{l_a, l_b} \sqrt{\binom{n}{l_a} \binom{m}{l_a} \binom{N-n}{l_b} \binom{N-m}{l_b}} |\alpha_n| |\alpha_m^*| \end{aligned} \quad (\text{B7})$$

Recalling that  $\beta_n^{l_a, l_b} = \sqrt{\binom{n}{l_a} \binom{N-n}{l_b}} \gamma_n^{l_a, l_b}$ , it is evident that the above equation is identical to Eq. (A4) obtained for the indistinguishable case. this completes the proof that the optimal estimation is indeed achievable using indistinguishable photons.

### Appendix C: Adaptive measurement schemes

Let us describe a general structure of adaptive measurement schemes performed on  $N$  subsystems. Let  $\{\Pi_{i_1}^{(1)}\}$  be a POVM performed on the first copy. Depending on the measurement result  $i_1$  a POVM  $\{\Pi_{i_2}^{(2)}(i_1)\}$  is performed on the second copy. In general, a POVM performed on the  $k$ -th copy  $\{\Pi_{i_k}^{(k)}(i_1, \dots, i_{k-1})\}$  depends on all previous measurement results. The adaptive measurement mathematically corresponds to POVM:

$$\Pi_{\mathbf{i}} = \Pi_{i_1, \dots, i_N} = \Pi_{i_1}^{(1)} \otimes \dots \otimes \Pi_{i_N}^{(N)}(i_1, \dots, i_{N-1}), \quad (\text{C1})$$

where  $\Pi_{\mathbf{i}}$  can be treated as a single global POVM with measurement results indexed by  $\mathbf{i}$ . This shows that, for distinguishable subsystems, optimization of estimation strategy over global POVMs covers also the case of adaptive measurements. Moreover, we have proved earlier in Appendix B, that the optimal phase estimation can be realized using indistinguishable subsystems. Therefore, the bounds derived in the paper, which assume a global POVM on indistinguishable photons, indeed hold also for all adaptive measurement strategies.

- 
- [1] J. J. . Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A, **54**, R4649 (1996).
  - [2] J. P. Dowling, Phys. Rev. A, **57**, 4736 (1998).
  - [3] D. W. Berry and H. M. Wiseman, Phys. Rev. Lett., **85**, 5098 (2000).
  - [4] V. Giovannetti, S. Lloyd, and L. Maccone, Science, **306**, 1330 (2004).
  - [5] M. W. Mitchell, J. S. Lundeen, and A. M. Steinberg, Nature, **429**, 161 (2004).
  - [6] P. Walther, J.-W. Pan, M. Aspelmeyer, R. Ursin, S. Gasparoni, and A. Zeilinger, Nature, **429**, 158 (2004).
  - [7] H. S. Eisenberg, J. F. Hodelin, G. Khoury, and D. Bouwmeester, Phys. Rev. Lett., **94**, 090502 (2005).
  - [8] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature, **450**, 393 (2007).
  - [9] T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki, and S. Takeuchi, Science, **316**, 726 (2007).
  - [10] M. Kacprowicz, R. Demkowicz-Dobrzanski, W. Wasilewski, and K. Banaszek, Nature Photonics, **4**, 357 (2010).
  - [11] J. Chwedenczuk, F. Piazza, and A. Smerzi, arxiv:1007.0703 (2010).
  - [12] M. G. A. Paris, Physics Letters A, **201**, 132 (1995).
  - [13] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Phys. Rev. Lett., **79**, 3865 (1997).
  - [14] M. A. Rubin and S. Kaushik, Phys. Rev. A, **75**, 053805 (2007).
  - [15] S. Olivares and M. G. A. Paris, Optics and Spectroscopy, **103**, 231 (2007).
  - [16] G. Gilbert, M. Hamrick, and Y. S. Weinstein, J. Opt. Soc. Am B, **25**, 1336 (2008).
  - [17] S. D. Huver, C. F. Wildfeuer, and J. P. Dowling, Phys. Rev. A, **78**, 063828 (2008).
  - [18] U. Dorner, R. Demkowicz-Dobrzanski, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Phys. Rev. Lett., **102**, 040403 (2009).
  - [19] R. Demkowicz-Dobrzanski, U. Dorner, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Phys. Rev. A, **80**, 013825 (2009).
  - [20] K. Banaszek, R. Demkowicz-Dobrzanski, and I. A. Walmsley, Nature Photonics, **3**, 673 (2009).
  - [21] T. Ono and H. F. Hofmann, Phys. Rev. A, **81**, 033819 (2010).
  - [22] R. Demkowicz-Dobrzański, Laser Physics, **20**, 1197 (2010).
  - [23] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, Phys. Rev. A, **80**, 052114 (2009).
  - [24] A. Hentschel and B. C. Sanders, Phys. Rev. Lett., **104**, 063603 (2010).
  - [25] C. W. Helstrom, *Quantum detection and estimation theory* (Academic press, 1976).
  - [26] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).
  - [27] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett., **72**, 3439 (1994).
  - [28] S. L. Braunstein, C. M. Caves, and G. J. Milburn, Annals of Physics, **247**, 135 (1996).
  - [29] In the absence of additional reference beams, a coherent superposition of states with different  $N$  becomes a statistical mixture, and the estimation cost for a mixture is always greater than the weighted cost for its constituents. Moreover, distinguishability of photons cannot decrease the cost (see Appendix B for the proof).
  - [30] In particular experimental setups,  $\eta_a$  and  $\eta_b$  may represent accumulated preparation, transmission

- and detection loss. E.g. in a Mach-Zehnder setup with detector efficiencies  $\xi$ , it is possible to formally commute detectors efficiency inside the interferometer, and set  $\eta'_a = \eta_a \xi$ ,  $\eta'_b = \eta_b \xi$ .
- [31] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Phys. Rev. A, **72**, 042338 (2005).
  - [32] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Rev. Mod. Phys., **79**, 555 (2007).
  - [33] M. D. Neusel and L. Smith, *Invariant Theory of Finite Groups* (American Mathematical Society, 2001).
  - [34] S. Knysh, V. N. Smelyanskiy, and G. A. Durkin, arXiv:1006.1645 (2010).